ISSN (Online): 2349-7084
GLOBAL IMPACT FACTOR 0.238
DIIF 0.876

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 9, SEPTEMBER 2015, PP 627-630

Building Confidential & Efficient Query Services in the Cloud with RASP Perturbation

[1] A Rebekah Johnson,[2] N.Parashuram, [3]Dr S.Prem Kumar
[1](M.Tech), CSE,
[2]Assistant Professor, Department of Computer Science and Engineering
[3]Professor & HOD, Department of computer science and engineering,
G.Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India.

Abstract:- In this paper Cloud computing infrastructures are popularly used by peoples now a days. By using cloud users can save their cost for query services. But some of the data owners are hesitate to put their data's in cloud because, sometimes the data may be hack when they use in cloud unless the confidentiality of data and secure query processing will be provided by the cloud provider.However, some data might be sensitive that the data owner does not want to move to the cloud unless the data confidentiality and query privacy are guaranteed. We propose the Random Space Encryption (RASP) approach that allows efficient range search with stronger attack resilience than existing efficiency-focused approaches. The random space perturbation (RASP) data perturbation method to provide secure and efficient range query and kNN query services for protected data in the cloud. The RASP data perturbation method combines order preserving encryption, dimensionality expansion, random noise injection, and random projection, to provide strong resilience to attacks on the perturbed data and queries. It also preserves multidimensional ranges, which allows existing indexing techniques to be applied to speedup range query processing. The kNN-R algorithm is designed to work with the RASP range query algorithm to process the kNN queries.

Keywords: RASP Method, query services in the cloud, privacy, range query, kNN query.

I.INTRODUCTION

With the wide deployment of public cloud computing infrastructures, using clouds to host data query services has become an appealing solution for the advantages on scalability and cost-saving. With the cloud infrastructures, the service owners can conveniently scale up or down the service and only pay for the hours of using the servers. While new approaches are needed to preserve data confidentiality and query privacy, the efficiency of query services and the benefits of using the clouds should also be preserved. It will not be meaningful to provide slow query services as a result of security and privacy assurance. It is also not practical for the data owner to use a significant amount of in-house resources, because the purpose of using cloud resources is to reduce the need of maintaining scalable in-house infrastructures. Therefore, there is an intricate relationship among the data confidentiality, query privacy, the quality of service, and the economics of using the cloud.[1] Here we summarize these requirements for constructing a practical query service in the cloud as the CPEL criteria: data confidentiality, query privacy, efficient query processing, and low in-house processing cost. Satisfying these requirements will dramatically increase the complexity of constructing query services in the cloud. Some related approaches have been developed to address some aspects of the problem. However, they do not satisfactorily address all of these aspects. For example, the cryptoindex and order preserving encryption (OPE) are vulnerable to the attacks. The enhanced cryptoindex approach puts heavy burden on the in-house infrastructure to improve the security and privacy. The New Casper approach uses cloaking boxes to protect data objects and queries, which affects the efficiency of query processing and the inhouse workload.

We propose the random space perturbation (RASP) approach to constructing practical range query and k-nearest- neighbor (kNN) query services in the cloud. The proposed approach will address all the four aspects of the CPEL criteria and aim to achieve a good balance on them. The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries.[1] The RASP perturbation is a unique combination of OPE, dimensionality expansion, random noise injection, and random projection, which provides strong confidentiality guarantee. We have carefully evaluated our approach with synthetic and real data sets. The results show its unique advantages on all aspects of the CPEL criteria. The RASP method and its combination provide confidentiality of data and this approach is mainly used to protect the multidimensional range of queries in secure manner, with indexing and efficient query processing. The range query is used in database for retrieving the stored data's. It will retrieve the records from the database where it can denote some value between upper and lower boundary. The kNN query denotes k-Nearest Neighbor query. K denotes positive integer and this query are used to find the value of nearest neighbor to k. The RASP perturbation embeds the multidimensional data into a secret higher dimensional space, en- hanced with random noise addition to protect the confidentiality of data.[2]

II.QUERY SERVICE

Query is mainly used to search. Queries are constructed by using structured query language. It is mainly used to retrieving the needed information from the database. Query services are the method for services that are exposed through an implementation of service provider. Here by using RASP, range query and kNN query in cloud provide secure, fast storing and retrieving process of encryption and decryption of a data from database. Range query is an important type of query for many data analytic tasks from simple aggregation to more sophisticated machine learning tasks. Let T be a table and Xi , Xj , and Xk be the real valued attributes in T, and a and b be some constants. Take the counting query for example. A typical range query looks like select count (*) from T where Xi ϵ [ai ,bi] and Xj ϵ (aj ,bj) and Xk=ak which calculates the number of records in the range defined by conditions on Xi , Xj , and Xk. Range queries may be applied to arbitrary number of attributes and conditions on these attributes combined with conditional operators "and"/"or." We call each part of the query condition that involves only one attribute as a simple condition. A simple condition like Xi ϵ[ai ,bi] can be described with two half space conditions Xi ≤ bi and −Xi ≤−ai . Without loss of generality, we will discuss how to process half-space conditions like Xi ≤ bi in this paper. A slight modification will extend the discussed algorithms to handle other conditions like Xi < bi and Xi = bi . kNN query is to find the closest k records to the query point, where the euclidean distance is often used to measure the proximity.

It is frequently used in locationbased services for searching the objects close to a query point, and also in machine learning algorithms such as hierarchical clustering and kNN classifier. A kNN query consists of the query point and the number of nearest neighbors, k.

III. SYSTEM ARCHITECTURE

Cloud computing infrastructures used to store large datasets and query services. The architecture shows two main parts in it. The data's can be stored in the cloud by data owners d=n (d, k) here d represents data, n represents normal form of data, k represents key value given by the data owner. This format will be saved in the cloud as encrypted form d=e (d, k) here e represents encryption.

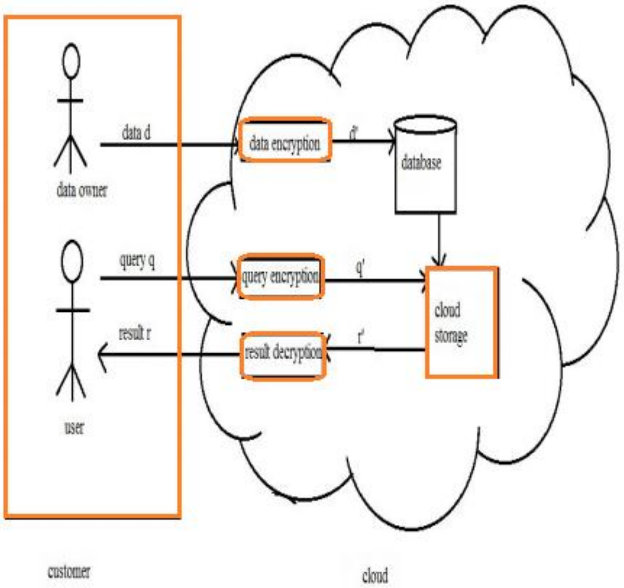

Fig. 1 System Architecture using RASP method

The above diagram shows two separate parties. They are customer who is the trusted party store their data in cloud and second are cloud provider who is storing the data in encrypted format. In the customer party it includes data and service owners, proxy server of in-house process and users. Here the owner can store their data in cloud while those data will encrypted in cloud and stored in the cloud database and also the data owner will provide key value by using that key value only cloud will encrypt the data by using random space perturbation method. The user will send query to retrieve the data from cloud, user can send range query and kNN query to get the data. In the cloud, the cloud provider has to host the user query services and have to protect the data stored in the cloud database. The basic procedure in the diagram is:

(1) The owner sends the data to store in cloud that data will be encrypted by using random space perturbation method and stored in cloud database.

(2) The user will send the range query or kNN query to retrieve the data that query will be encrypted and send to cloud storage.

(3) The cloud storage will send the data for that query after processing the query inside cloud storage and it will be decrypted and finally the data will send to the user.get the data.

In the cloud, the cloud provider has to host the user query services and have to protect the data stored in the cloud database. The basic procedure in the diagram is: (1) the owner sends the data to store in cloud that data will be encrypted by using random space perturbation method and stored in cloud database. (2) The user will send the range query or kNN query to retrieve the data that query will be encrypted and send to cloud storage. (3) the cloud storage will send the data for that query after processing the query inside cloud storage and it will be decrypted and finally the data will send to the user.

IV.SECURITY ANALYSIS

The security analysis in the architecture shows the following

- Users have been authorized by using the key value provided by the owner. So an authorized user is not being a malicious and only those users can send the queries for retrieving the data.
- The communication process between the user, owner and cloud and client system are well secured, the data and queries cannot be leaked from the cloud.
- RASP method is used to protect the query privacy and confidentiality of the data. Attacker Process: The main process of attacker is to hack the data from the database and they will try to find the perturbed data and they will try to find the queries.

V.MODULES

There Three modules are used. They are RASP, range query and kNN query.

5.1 RASP

RASP denotes Random Space Perturbation. It also combines OPE, random projection and random noise injection. Here OPE denotes Order Preserving Encryption is used for data that allows any comparison. And that comparison will be applied for the encrypted data; this will be done without decryption. Random projection is mainly used to process the high dimensional data into low dimensional data representations. It contains features like good scaling potential and good performances. Random noise injection is mainly used to adding noise to the input to get proper output when we compare it to the estimated power. The RASP method and its combination provide confidentiality of data and this approach is mainly used to protect the multidimensional range of queries in secure manner and also with indexing and efficient query processing will be done. RASP has some important features. In RASP the use of matrix multiplication does not protect the dimensional values so no need to suffer from the distribution based attack. RASP prevents the data that are perturbed from distance based attacks; it does not protect the distances that are occurred between the records. And also it won't protect more difficult structures it may be a matrix and other components. The range queries can be send to the RASP perturbed data and this range query describes open bounds in the multidimensional space. In random space perturbation, the word perturbation is used to do collapsing this process will happen according to the key value that is given by the owner. In this module the data owner have to register as owner and have to give owner name and key value. And then the user have register and get the key value and data owner name from the owner to do access in the cloud. Here user can submit their query as range query or kNN query and get their answer. We analyze and show the result with encrypted and also in decrypted format of the data for the query construct by the user.

5.2 RANGE QUERY

Range query is the query used to retrieve the data from the database. It will retrieve the data value that is between the upper bound and lower bound. The range query is not usual because user won't know in advance about the result for the query, how much entries will come as result for the query.

5.3 kNN QUERY kNN

Query represents k-Nearest Neighbor query. This query is mainly used to retrieve the nearest neighbor values of k. here k used to denote positive integer value. kNN algorithm is mainly used for classification and regression. In this it uses kNN-R algorithm to process the range query to kNN query. This algorithm consists of two methods. That is used to make interaction between the client and the server. The client will send the query to the server with initial upper bound and lower bound. This upper bound range has to be more than the k points and the lower bound range have to be less than the k points. The above process is used to give the inner range of the database by the server. With that inner range the client will calculate the outer range and send this outer range to the server. Then the server will search and find the records in the outer range from the database and send it to client and then the client will decrypt the record and find the top k files to provide the final result. This algorithm is used to find the compact inner square range for providing high precision and it has two difficult processes in it. They are to find the number of points that are present in the square range and updating of the boundary (i.e) upper bound and lower bound is difficult because range queries are well secured by using random space perturbation. The security of kNN query and range query is equal.

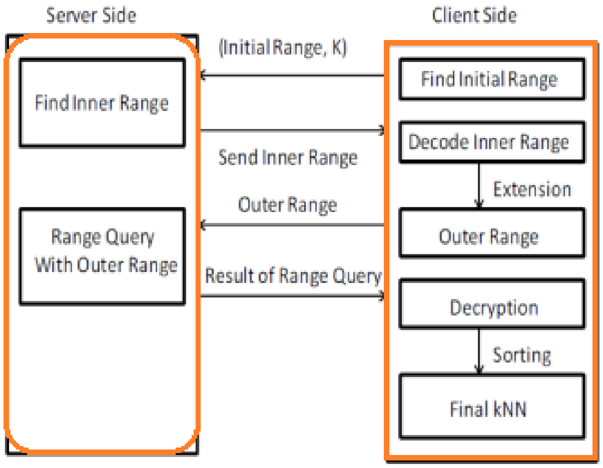

Fig. 2. Procedure of the KNN-R algorithm.

If the points are approximately uniformly distributed, we can estimate the precision of the returned result. With the uniform assumption, the number of points in an area is proportional to the size of the area. If the inner range contains m points, m > = k, the outer range contains q points, and the dimensionality is d, we can derive q = 2d=2m.

VI CONCLUSIONS

We proposed RASP method with range query and kNN query. This method mainly used to perturb the data given by the owner and saved in cloud storage it also combines random injection, order preserving encryption and random noise projection and also it has contains CPEL criteria in it. By using the range query and kNN query user can retrieve their data's in secured manner and the processing time of the query is minimized. And also we continue our studies to improve the effect of query.

REFERENCES

[1] Xu, H., Guo, S., and Chen, K. "Building confidential and efficient query services in the cloud with RASP data perturbation", IEEE Transactions on Knowledge and Data Engineering 26, 2 (2014).

[2] K. Chen, R. Kavuluru, and S. Guo, "RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases," Proc. ACM Conf. Data and Application Security and Privacy, pp. 249-260, 2011.

[3] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order Preserving Encryption for Numeric Data," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2004.

[4] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.K. Andy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing," technical report, Univ. of Berkerley, 2009.

[5] J. Bau and J.C. Mitchell, "Security Modeling and Analysis," IEEE Security and Privacy, vol. 9, no. 3, pp. 18- 25, May/June 2011.

[6]. M. L. Liu, G. Ghinita, C. S.Jensen, and P. Kalnis, "Enabling search services on outsourced private spatial data," The International Journal of on Very Large Data Base, vol. 19, no. 3, 2010.

[7]. M. F. Mokbel, C. yin Chow, and W. G. Aref, "The new casper: Query processing for location services without compromising privacy," in Proceedings of Very Large Databases Conference (VLDB), 2006, pp. 763–774.

[8]. M. Rudelson and R. Vershynin, "Smallest singular value of a random rectangular matrix," Communications on Pure and Applied Mathematics, vol. 62, pp. 1707–1739, 2009.

[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS), 2010.

[10]. E. Shi, J. Bethencourt, T.-H. H. Chan, D. Song, and A. Perrig, "Multi-dimensional range query over encrypted data," in IEEE Symposium on Security and Privacy, 2007.

[11]. P. Williams, R. Sion, and B. Carbunar, Building castles out of mud: Practical access pattern privacy and correctness on untrusted storage," in ACM Conference on Computer and Communications Security, 2008.

[12]. W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proceedings of ACM SIGMOD Conference. New York, NY, USA: ACM, 2009, pp. 139–152.